

# Тема урока

## Построение регрессионных моделей в Microsoft Excel

**Цель урока:** освоение способов построения по экспериментальным данным регрессионной модели средствами ЭТ.

# Основные этапы компьютерного моделирования

**1. Постановка задачи** - описание объекта моделирования в общем виде, определение конечного результата моделирования и имеющихся условий.

**2. Определение цели моделирования.** От выбранной цели зависит, какие характеристики объекта моделирования считать существенными, какие методы лучше подходят для решения данной задачи.

**3. Анализ объекта моделирования для выделения существенных свойств с точки зрения цели моделирования.**

**4. Формализация.** Формами представления данных могут быть: словесное описание, чертеж, таблица, формула, схема, алгоритм, компьютерная программа.

Конечной целью этого этапа является создание **информационной модели**.

**5. Разработка компьютерной модели для проведения эксперимента:**

Разнообразное программное обеспечение позволяет преобразовать исходную информационную модель в компьютерную и провести компьютерный эксперимент.

**6. Компьютерный эксперимент:**

а) исследование модели;

б) анализ полученных результатов на соответствие цели моделирования;

в) уточнение модели.

**Конечная цель моделирования** — принятие решения, которое должно быть выработано на основе всестороннего анализа результатов моделирования. Если результаты не соответствуют целям поставленной задачи, значит, на предыдущих этапах были допущены ошибки. Если такие ошибки выявлены, то требуется уточнение (корректировка) модели.

# Контрольные вопросы по табличному редактору Excel

## 1. Назначение ЭТ.

- a) для хранения и обработки числовых данных
- b) для хранения таблиц
- c) для обработки графической информации
- d) для представления текстовой информации в табличном виде.

## 2. Укажите форматы следующих данных:

12345, Мир, 14 апреля 1998г., =A2+17.

## 3. Приведите примеры абсолютных и относительных ссылок.

4. В ячейку D13 ввели формулу =C12/\$B\$4. Данную формулу скопировали в ячейку F15. Какая формула содержится в F15.

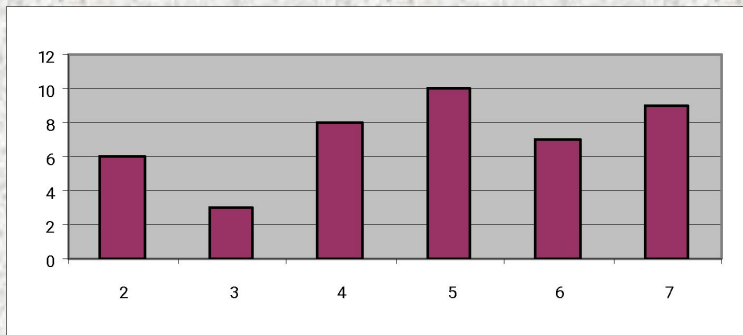
5. Сколько ячеек содержит блок D5:I9.

6. Какие категории функций, используемых в ЭТ, вы знаете.

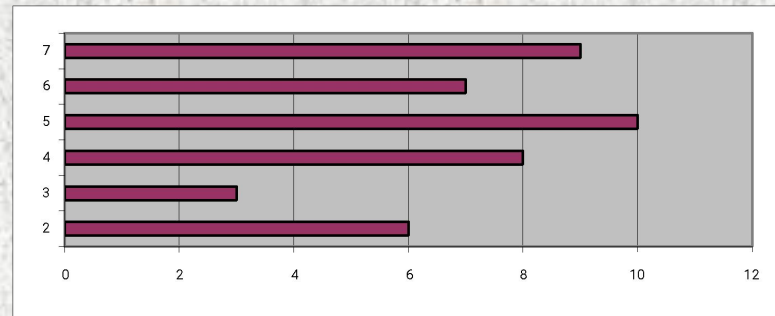
7. Какие статистические функции вы знаете.

# 8. Назовите типы приведенных ниже диаграмм.

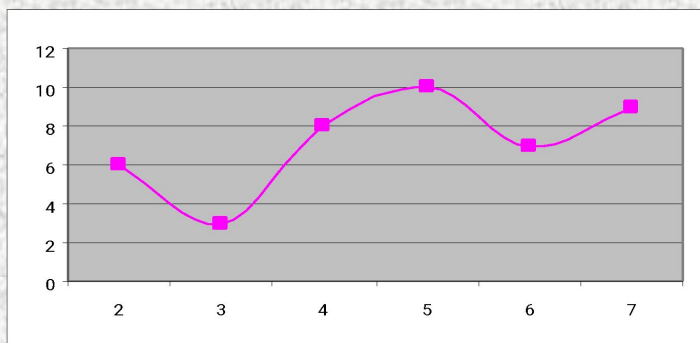
а)



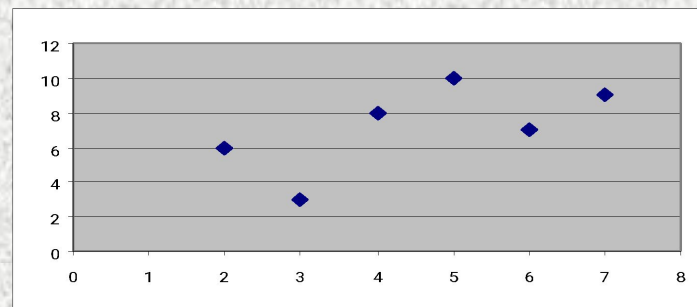
б)



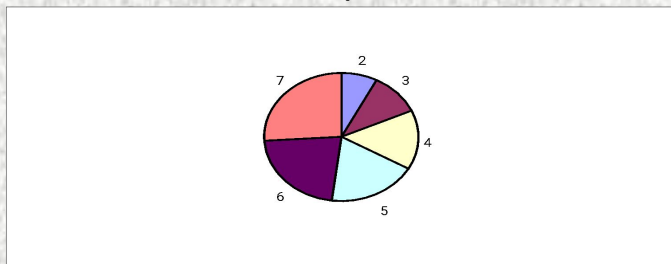
в)



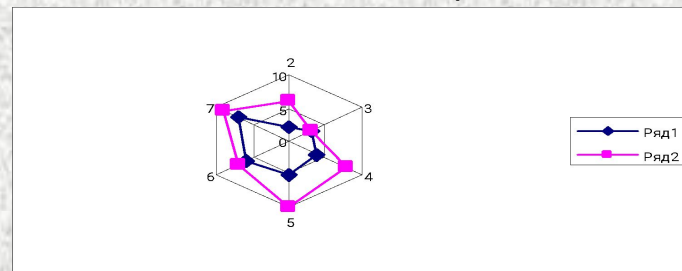
г)



д)



е)





# Математическая обработка статистических данных, результатов эксперимента

Зависимости между параметрами некоторого объекта, процесса, явления могут быть выражены с помощью математических формул. Но в некоторых случаях коэффициенты в этих формулах могут быть получены в результате статистической обработки экспериментальных данных.

**Статистика — это наука о сборе, измерении и анализе больших массивов данных.**

Статистические данные носят приближенный, усредненный характер, получаются путем многократных измерений. Статистические функции электронных таблиц позволяют обрабатывать статистические данные.

**Статистический анализ** данных широко используется:

- в народном хозяйстве при: анализе результатов деятельности предприятий и организаций;
- оценке состояния финансового, сырьевого и других рынков;
- анализе прибыльности инвестиционной деятельности;
- составлении краткосрочных планов и долгосрочных прогнозов;
- в научной деятельности для обработки результатов экспериментов, оценки их достоверности, проверки гипотез и пр.

Для обработки и представления результатов естественно-научного и математического эксперимента, экономических и экологических наблюдений, социальных опросов используются динамические (электронные) таблицы.

**Для определения статистической зависимости необходимо выполнить два шага:**

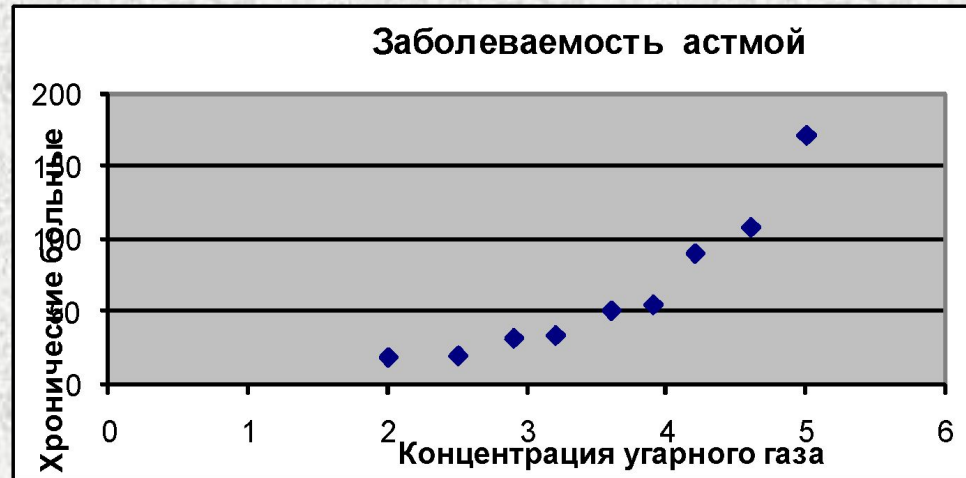
1. На основании физического смысла статистических данных выбрать вид функции таким образом, чтобы функция располагалась как можно ближе к экспериментальным данным.
2. Выбрать метод вычисления параметров функции, определяющих конкретный вид принятой зависимости.

**Полученная аналитическая зависимость называется регрессионной моделью.**

# Построение регрессионной модели.

Рассмотрим пример из медицинской статистики. Специалистами собраны сведения о средней концентрации угарного газа в атмосфере **C** и о заболеваемости астмой (число хронических больных на 1000 жителей) **P**. Полученные экспериментальные данные представим в табличной форме и в виде точечной диаграммы.

C мг/куб.м.	P бол./тыс.
2	19
2,5	20
2,9	32
3,2	34
3,6	51
3,9	55
4,2	90
4,6	108
5	171



Как теперь построить математическую модель данного явления? Нужно получить формулу зависимости **P** от **C**.

## Основные требования к искомой функции:

- она должна быть достаточно простой для использования ее в дальнейших вычислениях;
- график этой функции должен проходить вблизи экспериментальных точек так, чтобы отклонения этих точек от графика были минимальны и равномерны.

## Получение регрессионной модели происходит в два этапа:

- подбор вида функции;
- вычисление параметров функции.

Чаще всего выбор производится среди следующих функций:

- $y=ax+b$  линейная функция;
- $y=ax^2+bx+c$  квадратичная функция;
- $y=a\ln(x)+b$  логарифмическая функция;
- $y=ae^{bx}$  экспоненциальная функция;
- $y=ax^b$  степенная функция.

Во всех этих формулах  $x$  – аргумент,  $y$  – значение функции,  $a$ ,  $b$ ,  $c$  – параметры функций.

При выборе одной из функций нужно подобрать параметры так, чтобы функция располагалась как можно ближе к экспериментальным точкам.

Существует **метод наименьших квадратов (МНК)**.

Суть – искомая функция должна быть построена так, чтобы сумма квадратов отклонений  $y$ -координат всех экспериментальных точек от  $y$ -координат графика функции была бы минимальна.

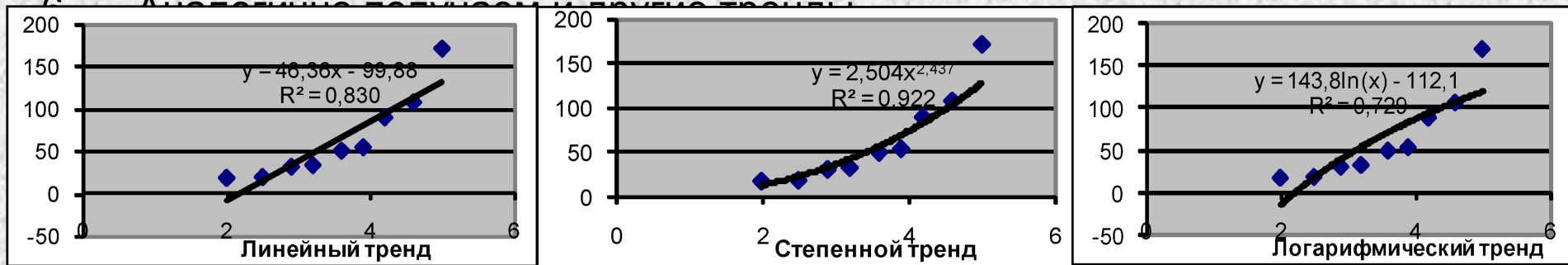
Графики регрессионной модели называются **трендами**. (Английское слово trend переводиться как общее направление или тенденция).



## Алгоритм получения с помощью MS Excel регрессионных моделей по МНК с построением тренда.

1. Ввести табличные данные.
2. Построить точечную диаграмму, где в качестве подписи к оси ОХ выбрать текст «Линейный тренд» (остальные надписи и легенду можно игнорировать).
3. Щелкнуть мышью по полю диаграммы; выполнить команду **Диаграмма – Добавить линию тренда**;
4. В открывшемся окне на закладке «Тип» выбрать «Линейный тренд»;
5. Перейти к закладке «Параметры» и установить галочки на флажках «показать уравнения на диаграмме» и «поместить на диаграмме величину достоверности аппроксимации  $R^2$ » и щелкнуть ОК.

С. Аналитические подписи к различным трендам



Величина  $R^2$  определяет, насколько удачной является полученная регрессионная модель. Из трех построенных моделей самая неудачная – **логарифмическая**, наиболее удачная – **степенная**.



## Практическое задание для самостоятельного выполнения на получение регрессионных зависимостей.

В представленной таблице приводится прогноз средней дневной температуры на последнюю неделю мая в различных городах европейской части России. Города упорядочены по алфавиту. Указана также географическая широта этих городов.

**Построить несколько вариантов регрессионных моделей (не менее трех), отражающих зависимость температуры от широты города. Выбрать наиболее подходящую регрессионную модель.**

Город	Широта, гр.с.ш.	Температура
Воронеж	51,5	16
Краснодар	45	24
Липецк	52,6	12
Новороссийск	44,8	25
Ростов-на-Дону	47,3	19
Рязань	54,5	11
Северодвинск	64,8	5
Череповец	59,4	7
Ярославль	57,7	10

## Домашнее задание:

По данным из следующей таблицы постройте с помощью MS Excel линейную, квадратичную, экспоненциальную и логарифмическую регрессионные модели.

***Запишите в тетрадь все уравнения регрессии и соотв. им величины достоверности аппроксимации. Сделайте вывод.***

X	2	4	6	8	10	12	14	16	18	20	22	24	26	28
Y	44	32	40	30	27	21	25	20	23	18	19	20	16	32

## **Практическое задание для самостоятельного выполнения на получение регрессионных зависимостей.**

В представленной таблице приводится прогноз средней дневной температуры на последнюю неделю мая в различных городах европейской части России. Города упорядочены по алфавиту. Указана также географическая широта этих городов.

**Построить несколько вариантов регрессионных моделей (не менее трех), отражающих зависимость температуры от широты города. Выбрать наиболее подходящую регрессионную модель.**

<b>Город</b>	<b>Широта, гр.с.ш.</b>	<b>Температура</b>
<b>Воронеж</b>	<b>51,5</b>	<b>16</b>
<b>Краснодар</b>	<b>45</b>	<b>24</b>
<b>Липецк</b>	<b>52,6</b>	<b>12</b>
<b>Новороссийск</b>	<b>44,8</b>	<b>25</b>
<b>Ростов-на-Дону</b>	<b>47,3</b>	<b>19</b>
<b>Рязань</b>	<b>54,5</b>	<b>11</b>
<b>Северодвинск</b>	<b>64,8</b>	<b>5</b>
<b>Череповец</b>	<b>59,4</b>	<b>7</b>
<b>Ярославль</b>	<b>57,7</b>	<b>10</b>



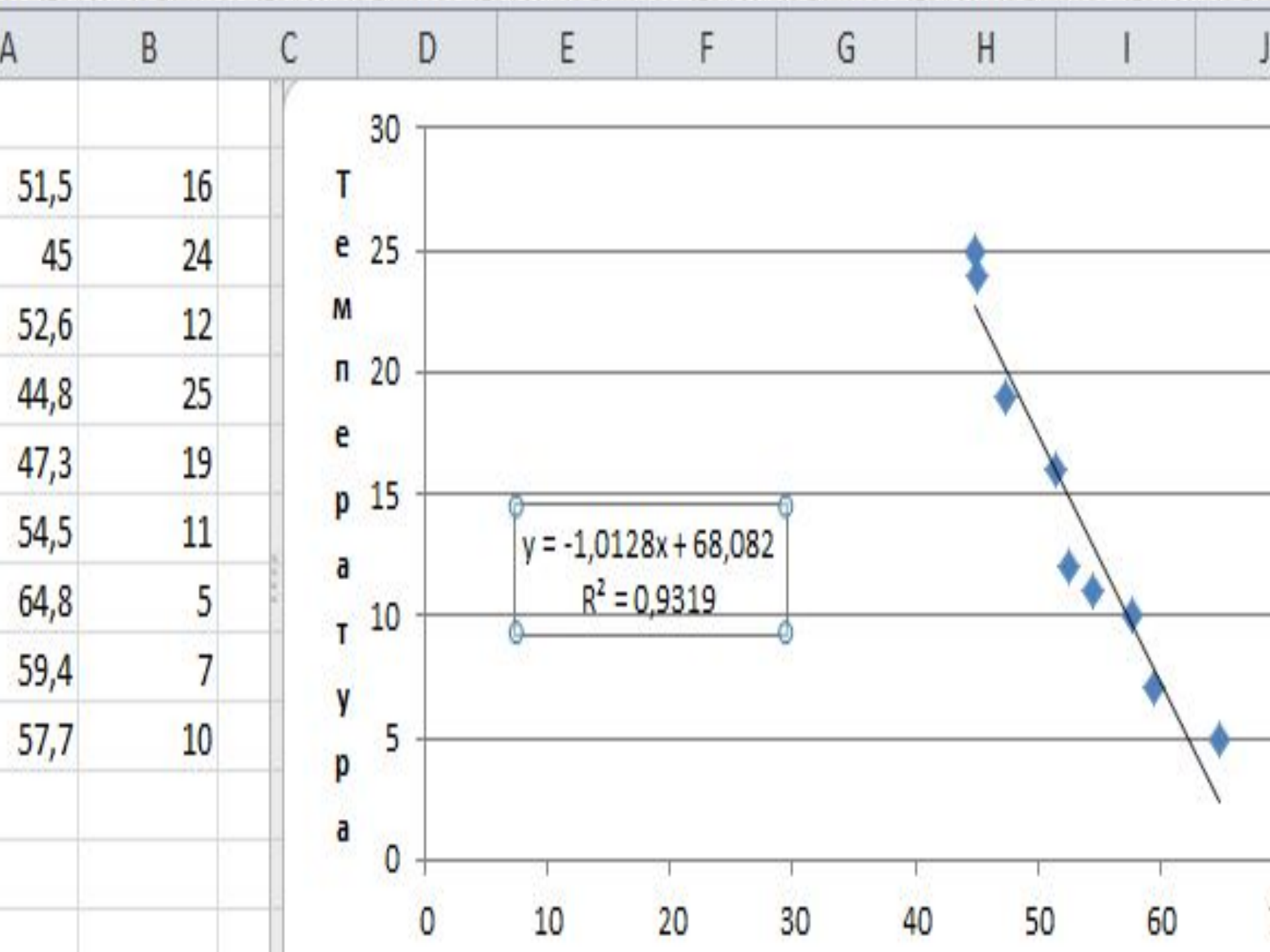
# Получение регрессионных моделей

Д/зад. В таблице приводится прогноз средней дневной температуры на последнюю неделю мая в городах России. Города упорядочены по алфавиту. Указана географическая широта этих городов.

Построить четыре варианта регрессионных моделей, отражающих зависимость температуры от широты города. Выбрать наиболее подходящую функцию.

*Записать все уравнения регрессии и соотв. им величины достоверности аппроксимации. Сделать вывод.*

<b>Город</b>	<b>Широта, гр.с. ш.</b>	<b>Температура</b>
<b>Воронеж</b>	<b>51,5</b>	<b>16</b>
<b>Краснодар</b>	<b>45</b>	<b>24</b>
<b>Липецк</b>	<b>52,6</b>	<b>12</b>
<b>Новороссийск</b>	<b>44,8</b>	<b>25</b>
<b>Ростов-на-Дону</b>	<b>47,3</b>	<b>19</b>
<b>Рязань</b>	<b>54,5</b>	<b>11</b>
<b>Северодвинск</b>	<b>64,8</b>	<b>5</b>





# Задание для самостоятельной работы

1. В таблице показана зависимость стоимости грузоперевозки от времени перевозки товара.

**Построить три варианта регрессионных моделей (экспоненциальную, полиномиальную и степенную), отражающих зависимость стоимости грузоперевозки от времени перевозки товара. Выбрать наиболее подходящую функцию.**

***Записать в тетрадь все уравнения регрессии и соотв. им величины достоверности аппроксимации. Сделать вывод.***

Пункт отправления	Пункт назначения	Стоимость (у. е.)	Время в пути
Москва	Пермь	100	70
Москва	Курск	30	10
Москва	Череповец	50	15
Пермь	Москва	100	69
Череповец	Пермь	140	80
Череповец	Москва	50	15
Череповец	Курск	100	80
Курск	Пермь	60	40
Курск	Москва	30	10
Курск	Череповец	100	80
Курск	Череповец	90	100

# Задание для самостоятельной работы

2. В таблице показана зависимость стоимости грузоперевозки от времени перевозки товара.

**Построить три варианта регрессионных моделей (линейную, логарифмическую и степенную), отражающих зависимость стоимости грузоперевозки от времени перевозки товара. Выбрать наиболее подходящую функцию.**

***Записать в тетрадь все уравнения регрессии и соотв. им величины достоверности аппроксимации. Сделать вывод.***



Пункт отправления	Пункт назначения	Стоимость (у. е.)	Время в пути
Москва	Пермь	100	70
Москва	Курск	30	10
Москва	Череповец	50	15
Пермь	Москва	100	69
Череповец	Пермь	140	80
Череповец	Москва	50	15
Череповец	Курск	100	80
Курск	Пермь	60	40
Курск	Москва	30	10
Курск	Череповец	100	80
Курск	Череповец	90	100